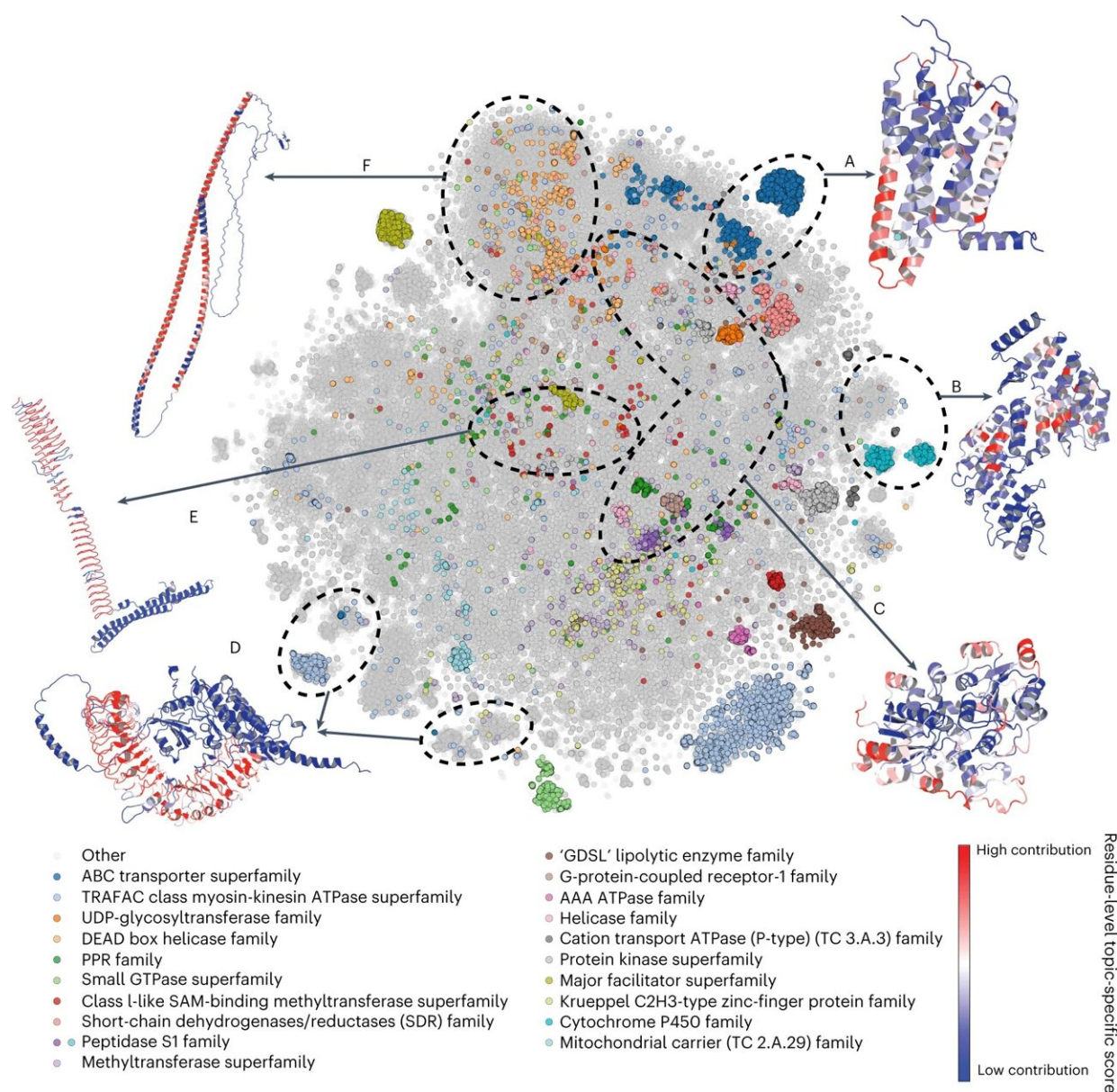


# 3D protein structure predictions made by AI could boost cancer research and drug discovery

November 8 2022



The space of characteristic structural elements in AF2 structural models for 21 species. Visualization of *t*-SNE dimensionality reduction analysis, in which structures with similar structural elements are placed closer together and the 20 most common superfamilies are colored. The axes corresponding to the *t*-SNE dimension 1 and *t*-SNE dimension 2 were omitted. Six shape-mer groups (that is, topics) discussed in the text, consisting of mainly AF2 proteins as opposed to PDB proteins, are labeled A–F, and a representative structure is depicted for each. Residues in the representative structures are colored according to their contribution to the topic under consideration—red residues have the highest contribution, and blue residues are specific to the example and not to the topic. Credit: *Nature Structural & Molecular Biology* (2022). DOI: 10.1038/s41594-022-00849-w

In a living being, proteins make up roughly everything: from the molecular machines running every cell's metabolism, to the tip of your hair. Encoded in the DNA, a protein may be represented as a thread of hundreds of individual molecules called amino acids, linked together. Depending on its particular amino acid combination, a protein folds in one way or another, resulting in a functional 3D shape. The shape makes the function, and with 20 different amino acids available, the possible combinations are countless.

Current genomic technologies make it very easy to know the amino acid sequence of a protein but knowing its 3D shape requires expensive and time-consuming experimental procedures, which are not always successful. For decades, researchers have tried to understand what makes a protein fold in a particular shape, to predict it from its amino acid sequence.

Alpha Fold 2 is a [neural network](#) developed by Deep Mind, a Google-owned Artificial Intelligence company, specifically trained to solve the

3D [structure of proteins](#) precisely from its amino acid sequence. Its accuracy impressed the [scientific community](#) a few years ago after its victories at the annual international contest on [protein structure](#) modeling CASP, when its team presented the full proteome for 11 different species, including humans.

To put all the data released by Alpha Fold 2 into context (over 300k models and growing), a community of independent researchers including Dr. Eduard Porta, head of the Cancer Immunogenetics group at the Josep Carreras Leukaemia Research Institute, compared the new structures made available to those currently available and concluded that Alpha Fold 2 contributed an extra 25% of high-quality protein structures to any given species. Their analysis has been recently published in *Nature Structural & Molecular Biology*.

The key role that many proteins play in disease, such as cancer, is already known, but the lack of a deep knowledge of their functioning at the molecular level prevents the development of specific strategies against them. The structural information of these proteins will help scientists to understand those proteins much better, to know what other molecules they may interact with inside the cell and to design new drugs, capable of interfering with their function when they are altered.

There are limitations, of course, to the capabilities of Alpha Fold 2. The community team found the algorithm has problems when trying to recreate protein complexes. Most proteins work together with other proteins to get a biological function done, so predicting how different proteins could stick together would be highly desirable. Another limitation identified is its inability to show the structure of mutated proteins, with altered [amino acids](#) on its sequence. Mutations often result in abnormal protein function and are the cause of many diseases like cancer.

Despite its limitations, however, the team recognizes the outstanding contribution of Alpha Fold 2 to the community, that will impact basic and biomedical research greatly in the coming years. Not only thanks to its direct contribution (thousands of new reliable 3D [protein](#) models), but by starting a new era of computational tools based on artificial intelligence able to yield results that no one can anticipate.

As a matter of fact, this era has already started and, recently, a team at Meta (formerly Facebook) has used a modified version of its natural language predictor to "autocomplete" proteins. This AI tool, called ESMFold, seems to be less accurate compared to its Google counterpart, but is 60 times faster and can overcome some of the identified Alpha Fold 2 limitations, such as handling mutated sequences.

All in all, as the authors of the publication admit, "the application of AlphaFold2 [and the coming tools] will have a transformative impact in life sciences."

**More information:** Mehmet Akdel et al, A structural biology community assessment of AlphaFold2 applications, *Nature Structural & Molecular Biology* (2022). [DOI: 10.1038/s41594-022-00849-w](https://doi.org/10.1038/s41594-022-00849-w)

Provided by Josep Carreras Leukaemia Research Institute

Citation: 3D protein structure predictions made by AI could boost cancer research and drug discovery (2022, November 8) retrieved 1 May 2024 from <https://phys.org/news/2022-11-3d-protein-ai-boost-cancer.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--